# The Semaphorin Gene Family

This application is a continuation of US Application Serial No. 09/060,610, filed April 15, 1998, which is a continuation of US Application Serial No. 08/835,268, filed April 8, 1997, now U.S. Patent No. 5,807,826, which is a division of US Application Serial No. 08/121,713, filed September 13, 1993, now U.S. Patent No. 5,639,856.

The research carried out in the subject application was supported in part by grants from the National Institutes of Health. The government may have rights in any patent issuing on this application.

## INTRODUCTION

### Technical Field

The technical field of this invention concerns peptides, polypeptides, and polynucleotides involved in nerve cell growth.

### Background

The specificity of the wiring of the nervous system -- the complex pattern of specific synaptic connections -- begins to unfold during development as the growing tips of neurons - the growth cones - traverse long distances to find their correct targets. Along their journey, they are confronted by and correctly navigate a series of choice points in a remarkably unerring way to ultimately contact and recognize their correct target.

The identification of growth cone guidance cues is to a large extent, the holy grail of neurobiology. These are the compounds that tell neurons when to grow, where to grow, and when to stop growing. The medical applications of such compounds and their antagonists are enormous and include modulating neuronal growth regenerative capacity, treating neurodegenerative disease, and mapping (e.g. diagnosing) genetic neurological defects.

Over decades of concentrated research, various hypotheses of chemo-attractants and repellant, labeled pathways, cell adhesion molecules, etc. have been evoked to explain guidance. Recently, several recent lines of experiments suggest repulsion may play an

important role in neuron guidance and two apparently unrelated factors ("Neurite Growth Inhibitor" and "Collapsin") capable of inhibiting or collapsing growth cones have been reported.

Relevant Literature

For a recent review of much of the literature in this field, see Goodman and Shatz (1993) Cell 72/Neuron 10, 77-98. A description of grasshopper fasciclin IV (now called G-Semaphorin I) appears in Kolodkin et al. (1992) Neuron 9, 831-845. Recent reports on Collapsin and Neurite Growth Inhibitor include Raper and Kapfhammer (1990) Neuron 4, 21-29, an abstract presented by Raper at the GIBCO-BRL Symposium on "Genes and Development/Function of Brain" on July 26, 1993 and Schwab and Caroni (1988) J Neurosci 8, 2381 and Schnell and Schwab (1990) Nature 343, 269, respectively.

## SUMMARY OF THE INVENTION

A novel class of proteins, semaphorins, nucleic acids encoding semaphorins, and methods of using semaphorins and semaphorin-encoding nucleic acids are disclosed. Semaphorins include the first known family of human proteins which function as growth cone inhibitors and a family of proteins involved in viral, particularly pox viral, pathogenesis and oncogenesis. Families of semaphorin-specific receptors, including receptors found on nerve growth cones and immune cells are also disclosed.

The invention provides agents, including semaphorin peptides, which specifically bind semaphorin receptors and agents, including semaphorin receptor peptides, which specifically bind semaphorins. These agents provide potent modulators of nerve cell growth, immune responsiveness and viral pathogenesis and find use in the treatment and diagnosis of neurological disease and neuro-regeneration, immune modulation including hypersensitivity and graft-rejection, and diagnosis and treatment of viral and oncological infection/diseases.

Semaphorins, semaphorin receptors, semaphorin-encoding nucleic acids, and unique portions thereof also find use variously in screening chemical libraries for regulators of semaphorin or semaphorin receptor-mediated cell activity, in genetic mapping, as probes for related genes, as diagnostic reagents for genetic neurological, immunological and oncological disease and in the production of specific cellular and animal systems for the development of neurological, immunological, oncological and viral disease therapy.

## DESCRIPTION OF SPECIFIC EMBODIMENTS

The present invention discloses novel families of proteins important in nerve and immune cell function: the semaphorins and the semaphorin receptors. The invention provides agents, including semaphorin peptides, which specifically bind semaphorin receptors and

agents, including semaphorin receptor peptides, which specifically bind semaphorins. These agents find a wide variety of clinical, therapeutic and research uses, especially agents which modulate nerve and/or immune cell function by specifically mimicing or interfering with semaphorin-receptor binding. For example, selected semaphorin peptides shown to act as semaphorin receptor antagonists are effective by competitively inhibiting native semaphorin association with cellular receptors. Thus, depending on the targeted receptor, these agents can be used to block semaphorin mediated neural cell growth cone repulsion or contact inhibition. Such agents find broad clinical application where nerve cell growth is indicated, e.g. traumatic injury to nerve cells, neurodegenerative disease, etc. A wide variety of semaphorin- and semaphorin receptor-specific binding agents and methods for identifying, making and using the same are described below.

Binding agents of particular interest are semaphorin peptides which specifically bind and antagonize a semaphorin receptor and semaphorin receptor peptides which specifically bind a semaphorin and prevent binding to a native receptor. While exemplified primarily with semaphorin peptides, much of the following description applies analogously to semaphorin receptor peptides.

The semaphorin peptides of the invention comprise a unique portion of a semaphorin and have semaphorin binding specificity. A "unique portion" of a semaphorin has an amino acid sequence unique to that disclosed in that it is not found in any previously known protein. Thus a unique portion has an amino acid sequence length at least long enough to define a novel peptide. Unique semaphorin portions are found to vary from about 5 to about 25 residues, preferably from 5 to 10 residues in length, depending on the particular amino acid sequence. Unique semaphorin portions are readily identified by comparing the subject semaphorin portion sequences with known peptide/protein sequence data bases. Preferred unique portions derive from the semaphorin domains (which exclude the Ig-like, intracellular and transmembrane domains as well as the signal sequences) of the disclosed semaphorin sequences, especially regions that bind the semaphorin receptor, especially that of the human varieties. Preferred semaphorin receptor unique portions derive from the semaphorin binding domains, especially regions with residues which contact the semaphorin ligand, especially that of the human varieties. Particular preferred peptides are further described herein.

The subject peptides may be free or coupled to other atoms or molecules. Frequently the peptides are present as a portion of a larger polypeptide comprising the subject peptide where the remainder of the polypeptide need not be semaphorin- or semaphorin receptor-derived. Alternatively, the subject peptide may be present as a portion of a "substantially full-length" semaphorin domain or semaphorin receptor sequence which comprises or encodes at least about 200, preferably at least about 250, more preferably at least about 300 amino acids of a disclosed semaphorin/receptor sequence. Thus the invention also provides

polypeptides comprising a sequence substantially similar to that of a substantially full-length semaphorin domain or a semaphorin receptor. "Substantially similar" sequences share at least about 40%, more preferably at least about 60%, and most preferably at least about 80% sequence identity. Where the sequences diverge, the differences are generally point insertions/deletions or conservative substitutions, i.e. a cysteine/threonine or serine substitution, an acidic/acidic or hydrophobic/hydrophobic amino acid substitution, etc.

The subject semaphorin peptides/polypeptides are "isolated", meaning unaccompanied by at least some of the material with which they are associated in their natural state. Generally, an isolated peptide/polypeptide constitutes at least about 1%, preferably at least about 10%, and more preferably at least about 50% by weight of the total peptide/protein in a given sample. By pure peptide/polypeptide is intended at least about 90%, preferably at least 95%, and more preferably at least about 99% by weight of total peptide/protein. Included in the subject peptide/polypeptide weight are any atoms, molecules, groups, or polymers covalently coupled to the subject semaphorin/receptor peptide/polypeptide, especially peptides, proteins, detectable labels, glycosylations, phosphorylations, etc.

The subject peptides/polypeptides may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample and to what, if anything, the peptide/polypeptide is covalently linked. Purification methods include electrophoretic, molecular, immunological and chromatographic techniques, especially affinity chromatography and RP-HPLC in the case peptides. For general guidance in suitable purification techniques, see Scopes, R., Protein Purification, Springer-Verlag, NY (1982).

The subject peptides/polypeptides generally comprise naturally occurring amino acids but D-amino acids or amino acid mimetics coupled by peptide bonds or peptide bond mimetics may also be used. Amino acid mimetics are other than naturally occurring amino acids that conformationally mimic the amino acid for the purpose of the requisite semaphorin/receptor binding specificity. Suitable mimetics are known to those of ordinary skill in the art and include $\beta$-$\gamma$-$\delta$ amino and imino acids, cyclohexylalanine, adamantylacetic acid, etc., modifications of the amide nitrogen, the $\alpha$-carbon, amide carbonyl, backbone modifications, etc. See, generally, Morgan and Gainor (1989) Ann. Repts. Med. Chem 24, 243-252; Spatola (1983) Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol VII (Weinstein) and Cho et. al (1993) Science 261, 1303-1305 for the synthesis and screening of oligocarbamates.

The subject semaphorin peptides/polypeptides have a "semaphorin binding specificity" meaning that the subject peptide/polypeptide retains a molecular conformation specific to one or more of the disclosed semaphorins and specifically recognizable by a semaphorin-specific receptor, antibody, etc. As such, a semaphorin binding specificity may

be provided by a semaphorin-specific immunological epitope, lectin binding site, etc., and preferably, a receptor binding site. Analogously, the semaphorin receptor peptides/polypeptides have a "semaphorin receptor binding specificity" meaning that these peptides/polypeptides retain a molecular conformation specific to one or more of the disclosed semaphorin receptors and specifically recognizable by a semaphorin, a receptor-specific antibody, etc.

"Specific binding" is empirically determined by contacting, for example a semaphorin-derived peptide with a mixture of components and identifying those components that preferentially bind the semaphorin. Specific binding is most conveniently shown by competition with labeled ligand using recombinant semaphorin peptide either in vitro or in cellular expression systems as disclosed herein. Generally, specific binding of the subject semaphorin has binding affinity of $10^{-6}$M, preferably $10^{-8}$M, more preferably $10^{-10}$M, under in vitro conditions as exemplified below.

The peptides/polypeptides may be modified or joined to other compounds using physical, chemical, and molecular techniques disclosed or cited herein or otherwise known to those skilled in the relevant art to affect their semaphorin binding specificity or other properties such as solubility, membrane transportability, stability, binding specificity and affinity, chemical reactivity, toxicity, bioavailability, localization, detectability, in vivo half-life, etc. as assayed by methods disclosed herein or otherwise known to those of ordinary skill in the art. For example, point mutations are introduced by site directed mutagenesis of nucleotides in the DNA encoding the disclosed semaphorin polypeptides or in the course of in vitro peptide synthesis.

Other modifications to further modulate binding specificity/affinity include chemical/enzymatic intervention (e.g. fatty acid-acylation, proteolysis, glycosylation) and especially where the peptide/polypeptide is integrated into a larger polypeptide, selection of a particular expression host, etc. In particular, many of the disclosed semaphorin peptides contain serine and threonine residues which are phosphorylated or dephosphorylated. See e.g. methods disclosed in Roberts et al. (1991) Science 253, 1022-1026 and in Wegner et al. (1992) Science 256, 370-373. Amino and/or carboxyl termini may be functionalized e.g., for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like. Many of the disclosed semaphorin peptides/polypeptides also contain glycosylation sites and patterns which may disrupted or modified, e.g. by enzymes like glycosidases or used to purify/identify the receptor, e.g. with lectins. For instance, N or O-linked glycosylation sites of the disclosed semaphorin peptides may be deleted or substituted for by another basic amino acid such as Lys or His for N-linked glycosylation alterations, or deletions or polar substitutions are introduced at Ser and Thr residues for modulating O-linked glycosylation. Glycosylation variants are also produced by selecting

appropriate host cells, e.g. yeast, insect, or various mammalian cells, or by in vitro methods such as neuraminidase digestion. Useful expression systems include COS-7, 293, BHK, CHO, TM4, CV1, VERO-76, HELA, MDCK, BRL 3A, W138, Hep G2, MMT 060562, TRI cells, baculovirus systems, for examples. Other covalent modifications of the disclosed semaphorin peptides/polypeptides may be introduced by reacting the targeted amino acid residues with an organic derivatizing (e.g. methyl-3-[(p-azido-phenyl)dithio] propioimidate) or crosslinking agent (e.g. 1,1-bis(diazoacetyl)-2-phenylethane) capable of reacting with selected side chains or termini. For therapeutic and diagnostic localization, semaphorins and peptides thereof may be labeled directly (radioisotopes, fluorescers, etc.) or indirectly with an agent capable of providing a detectable signal, for example, a heart muscle kinase labeling site.

The following are 14 classes of preferred semaphorin peptides where bracketed positions may be occupied by any one of the residues contained in the brackets and "X" signifies that the position may be occupied by any one of the 20 naturally encoded amino acids (see, Table 1). These enumerated peptides maintain highly conserved structures which provide important semaphorin binding specificities;

(a) [AspGlu]Cys[GlnLysArgAlaAsn]Asn[TyrPheVal]Ile (SEQ ID NO:1)
Cys[GlnLysArgAlaAsn]Asn[TyrPheVal]Ile[ArgLysGlnThr] (SEQ ID NO:2)
(b)CysGlyThr[AsnGly][AlaSerAsn][TyrPheHisGly][LysArgHisAsnGln] (SEQ ID NO:3)
CysGlyThr[AsnGly][AlaSerAsn]XaaXaaPro (SEQ ID NO:4)
CysGlyThr[AsnGly]XaaXaaXaaProXaa[CysAsp] (SEQ ID NO:5)
CysGlyThrXaaXaaXaaXaaProXaa[CysAsp]XaaXaa[TyrIle] (SEQ ID NO:6)
(c)[ArgIleGlnVal][GlyAla][LeuValLys][CysSer]Pro[PheTyr][AspAsn] (SEQ ID NO:7)
[CysSer]Pro[PheTyr][AspAsn]Pro[AspGluArgLys][HisLeuAsp] (SEQ ID NO:8)
GlyXaa[GlyAla]Xaa[CysSer]ProTyr[AspAsn]Pro (SEQ ID NO:9)
(d) Leu[PheTyr]Ser[GlyAla]Thr[ValAsnAla]Ala (SEQ ID NO:10)
Leu[PheTyr]SerXaaThrXaaAla[AspGlu][PheTyr] (SEQ ID NO:11)
[PheTyr]Ser[GlyAla]Thr[ValAsnAla]Ala[AspGlu][PheTyr] (SEQ ID NO:12)
(e) Leu[AsnAsp][AlaLys]ProAsnPheVal (SEQ ID NO:13)
(f) PhePhePheArgGlu (SEQ ID NO:14)
PhePhe[PheTyr]ArgGlu[ThrAsn] (SEQ ID NO:15)
PhePheArgGlu[ThrAsn]Ala (SEQ ID NO:16)
Phe[PheTyr]ArgGlu[ThrAsn]Ala (SEQ ID NO:17)
TyrPhePhe[PheTyr]ArgGlu (SEQ ID NO:18)
[PheTyr]PhePhe[PheTyr]ArgGlu (SEQ ID NO:19)
[PheTyr][PheTyr][PheTyr]ArgGlu[ThrAsn]Ala (SEQ ID NO:20)

[IleVal][PheTyr]Phe[PheTyr][PheTyr]ArgGlu (SEQ ID NO:21)
Asp[LysPheTyr]Val[PheTyr][PheTyrIleLeu][PheTyrIleLeu][PheTyr] (SEQ ID NO:22)
[ValIle][PheTyr][PheTyrIleLeu][PheTyrIleLeu]Phe[ArgThr]Xaa[ThrAsn](SEQID NO:23)
[ValIle][PheTyr][PheTyrIleLeu][PheTyrIleLeu][PheTyr][ArgThr][GluAspVal][ThrAsn]
(SEQ ID NO:24)

(g) Glu[PheTyr]IleAsn[CysSer]GlyLys (SEQ ID NO:25)
[PheTyr]IleAsnCysGlyLys[AlaValIle] (SEQ ID NO:26)

(h) Arg[ValIle][AlaGly][ArgGln][ValIle]CysLys (SEQ ID NO:27)
Arg[ValIle]Xaa[ArgGln][ValIle]CysXaaXaaAsp (SEQ ID NO:28)
GlyLys[ValAlaIle]XaaXaaXaaArg[ValAlaIle]XaaXaaXaaCysLys (SEQ ID NO:29)

(i)[ArgLysAsn]Trp[ThrAlaSer][ThrAlaSer][PheTyrLeu]Leu[LysArg] (SEQ ID NO:30)
[PheTyr]Leu[LysArg][AlaSer]ArgLeu[AsnIle]Cys (SEQ ID NO:31)
[AsnIle]CysSer[IleVal][ProSer]Gly (SEQ ID NO:32)
Trp[ThrAlaSer][ThrAlaSer][PheTyrLeu]LeuLys[AlaSerValIleLeu]XaaLeu (SEQ ID NO:33)
Trp[ThrAlaSer][ThrAlaSer]XaaLeuLysXaaXaaLeuXaaCys (SEQ ID NO:34)
TrpXaa[ThrSer]XaaLeuLysXaaXaaLeuXaaCys (SEQ ID NO:35)

(j) [PheTyr][PheTyr][AsnAsp]GluIleGlnSer (SEQ ID NO:36)
[PheTyr]Pro[PheTyr][PheTyr][PheTyr][AsnAsp]Glu (SEQ ID NO:37)

(k) GlySerAla[ValIleLeu]CysXaa[PheTyr] (SEQ ID NO:38)
SerAla[ValIleLeu]CysXaa[PheTyr]XaaMet (SEQ ID NO:39)

(l) AsnSer[AsnAla]TrpLeu[ProAla]Val (SEQ ID NO:40)

(m) [ValLeuIle]Pro[GluAspTyrSerPhe]ProArgProGly (SEQ ID NO:41)
[ValLeuIle]ProXaaPro[ArgAla]ProGlyXaaCys (SEQ ID NO:42)
Pro[GluAspTyrSerPhe]ProArgProGly[ThrGlnSer]Cys (SEQ ID NO:43)

(n) AspPro[HisPheTyr]Cys[AlaGly]Trp (SEQ ID NO:44)
Pro[HisPheTyr]Cys[AlaGly]TrpAsp (SEQ ID NO:45)
AspProXaaCys[AlaGly]TrpAsp (SEQ ID NO:46)
CysXaaXaaXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:47)
CysXaaXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:48)
CysXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:49)
CysXaaXaaCysXaaXaaXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:50)
CysXaaXaaCysXaaXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:51)
CysXaaXaaCysXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:52).

The following peptides represent particularly preferred members of each class:

(a) AspCysGlnAsnTyrIle (SEQ ID NO:67)

(b) CysGlyThr[AsnGly][AlaSer]XaaXaaPro (SEQ ID NO:68)
(c) GlyXaa[SerCys]ProTyrAspPro (SEQ ID NO:69)
(d) LeuTyrSerGlyThr[ValAsnAla]Ala (SEQ ID NO:70)
(e) LeuAsnAlaProAsnPheVal (SEQ ID NO:71)
(f) [PheTyr]PhePhe[PheTyr]ArgGlu (SEQ ID NO:19)
(g) Glu[PheTyr]IleAsn[CysSer]GlyLys (SEQ ID NO:25)
(h) Arg[ValIle]AlaArgValCysLys (SEQ ID NO:72)
(i) Trp[ThrAla][ThrSer][PheTyr]LeuLys[AlaSer]ArgLeu (SEQ ID NO:73)
(j) ProPheTyrPhe[AsnAsp]GluIleGlnSer (SEQ ID NO:74)
(k) GlySerAlaValCysXaa[PheTyr] (SEQ ID NO:75)
(l) AsnSerAsnTrpLeu[ProAla]Val (SEQ ID NO:76)
(m) Pro[GluAsp]ProArgProGly[ThrGlnSer]Cys (SEQ ID NO:77)
(n) AspProTyrCys[AlaGly]TrpAsp (SEQ ID NO:78).

The following 14 classes are preferred peptides which exclude semaphorin peptides encoded in open reading frames of Variola major or Vaccinia viruses.

(a) [AspGlu]Cys[GlnLysArgAlaAsn]Asn[TyrPheVal]Ile (SEQ ID NO:01)
Cys[GlnLysArgAlaAsn]Asn[TyrPheVal]Ile[ArgLysGlnThr] (SEQ ID NO:02)
(b) CysGlyThr[AsnGly][AlaSer][TyrPheHisGly][LysArgHisAsnGln] (SEQ ID NO:79)
CysGlyThr[AsnGly][AlaSerAsn][TyrPheHis][LysArgHisAsnGln] (SEQ ID NO:80)
CysGlyThr[AsnGly][AlaSer]XaaXaaPro (SEQ ID NO:81)
(c)[ArgIleGlnVal][GlyAla][LeuValLys][CysSer]Pro[PheTyr][AspAsn] (SEQ ID NO:07)
[CysSer]Pro[PheTyr][AspAsn]Pro[AspGluArgLys][HisLeuAsp] (SEQ ID NO:08)
GlyXaa[GlyAla]Xaa[CysSer]ProTyr[AspAsn]Pro (SEQ ID NO:09)
(d) Leu[PheTyr]Ser[GlyAla]Thr[ValAsnAla]Ala (SEQ ID NO:10)
Leu[PheTyr]SerXaaThrXaaAla[AspGlu][PheTyr] (SEQ ID NO:11)
[PheTyr]Ser[GlyAla]Thr[ValAsnAla]Ala[AspGlu][PheTyr] (SEQ ID NO:12)
(e) Leu[AsnAsp][AlaLys]ProAsnPheVal (SEQ ID NO:13)
(f) PhePhePheArgGlu (SEQ ID NO:14)
PhePhe[PheTyr]ArgGlu[ThrAsn] (SEQ ID NO:15)
PhePheArgGlu[ThrAsn]Ala (SEQ ID NO:16)
Phe[PheTyr]ArgGlu[ThrAsn]Ala (SEQ ID NO:17)
TyrPhePhe[PheTyr]ArgGlu (SEQ ID NO:18)
[PheTyr]PhePhe[PheTyr]ArgGlu (SEQ ID NO:19)
[PheTyr][PheTyr][PheTyr]ArgGlu[ThrAsn]Ala (SEQ ID NO:20)
[IleVal][PheTyr]Phe[PheTyr][PheTyr]ArgGlu (SEQ ID NO:21)

Asp[LysPheTyr]Val[PheTyr][PheTyrLeu][PheTyrIleLeu][PheTyr] (SEQ ID NO:22)
Asp[LysPheTyr]Val[PheTyr][PheTyrIleLeu][PheTyrIle][PheTyr] (SEQ ID NO:82)
[ValIle][PheTyr][PheTyrLeu][PheTyrIleLeu]Phe[ArgThr]Xaa[ThrAsn] (SEQ ID NO:83)
[ValIle][PheTyr][PheTyrIleLeu][PheTyrIle]Phe[ArgThr]Xaa[ThrAsn] (SEQ ID NO:84)
[ValIle][PheTyr][PheTyrIleLeu][PheTyrIleLeu]PheArgXaa[ThrAsn] (SEQ ID NO:85)
[ValIle][PheTyr][PheTyrLeu][PheTyrIleLeu][PheTyr][ArgThr][GluAspVal][ThrAsn] (SEQ ID NO:86)

(g) Glu[PheTyr]IleAsn[CysSer]GlyLys (SEQ ID NO:25)
[PheTyr]IleAsnCysGlyLys[AlaValIle] (SEQ ID NO:26)

(h) Arg[ValIle][AlaGly][ArgGln][ValIle]CysLys (SEQ ID NO:27)
Arg[ValIle]Xaa[ArgGln][ValIle]CysXaaXaaAsp (SEQ ID NO:28)
GlyLys[ValAlaIle]XaaXaaXaaArg[ValAlaIle]XaaXaaXaaCysLys (SEQ ID NO:29)

(i) [ArgLysAsn]Trp[ThrAla][ThrAlaSer][PheTyrLeu]Leu[LysArg] (SEQ ID NO:87)
[PheTyr]Leu[LysArg][AlaSer]ArgLeu[AsnIle]Cys (SEQ ID NO:31)
[AsnIle]CysSer[IleVal][ProSer]Gly (SEQ ID NO:32)
Trp[ThrAla][ThrAlaSer][PheTyrLeu]LeuLys[AlaSerValIleLeu]XaaLeu (SEQ ID NO:88)
Trp[ThrAlaSer][ThrAlaSer][PheTyrLeu]LeuLys[AlaSerIleLeu]XaaLeu (SEQ ID NO:89)
Trp[ThrAla][ThrAlaSer]XaaLeuLysXaaXaaLeuXaaCys (SEQ ID NO:90)

(j) [PheTyr][PheTyr][AsnAsp]GluIleGlnSer (SEQ ID NO:36)
[PheTyr]Pro[PheTyr][PheTyr][PheTyr][AsnAsp]Glu (SEQ ID NO:37)

(k) GlySerAla[ValIleLeu]CysXaa[PheTyr] (SEQ ID NO:38)
SerAla[ValIle]CysXaa[PheTyr]XaaMet (SEQ ID NO:39)

(l) AsnSer[AsnAla]TrpLeu[ProAla]Val (SEQ ID NO:40)

(m) [ValLeuIle]Pro[GluAspTyrSerPhe]ProArgProGly (SEQ ID NO:41)
[ValLeuIle]ProXaaProArgProGlyXaaCys (SEQ ID NO:91)
Pro[GluAspTyrSerPhe]ProArgProGly[ThrGlnSer]Cys (SEQ ID NO:43)

(n) AspPro[HisPheTyr]Cys[AlaGly]Trp (SEQ ID NO:44)
Pro[HisPheTyr]Cys[AlaGly]TrpAsp (SEQ ID NO:45)
AspProXaaCys[AlaGly]TrpAsp (SEQ ID NO:46)
CysXaaXaaXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:47)
CysXaaXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:48)
CysXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:49)
CysXaaXaaCysXaaXaaXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:50)
CysXaaXaaCysXaaXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:51)
CysXaaXaaCysXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:52).

The following 2 classes are preferred peptides which exclude semaphorin peptides

encoded in open reading frames of Variola major or Vaccinia viruses Grasshopper Semaphorin I.

(f) TyrPhePhe[PheTyr]ArgGlu (SEQ ID NO:18)
Asp[LysTyr]Val[PheTyr][PheTyrLeu][PheTyrIleLeu][PheTyr] (SEQ ID NO:92)
Asp[LysTyr]Val[PheTyr][PheTyrIleLeu][PheTyrIle][PheTyr] (SEQ ID NO:93)
[ValIle]Tyr[PheTyrLeu][PheTyrIleLeu]Phe[ArgThr]Xaa[ThrAsn] (SEQ ID NO:94)
[ValIle]Tyr[PheTyrIleLeu][PheTyrIle]Phe[ArgThr]Xaa[ThrAsn] (SEQ ID NO:95)
[ValIle]Tyr[PheTyrIleLeu][PheTyrIleLeu]PheArgXaa[ThrAsn] (SEQ ID NO:96)
Val[PheTyr][PheTyrLeu][PheTyrIleLeu][PheTyr][ArgThr][GluAspVal][ThrAsn] (SEQ ID NO:97)
Val[PheTyr][PheTyrIleLeu][PheTyrIle][PheTyr][ArgThr][GluAspVal][ThrAsn] (SEQ ID NO:98)
Val[PheTyr][PheTyrIleLeu][PheTyrIleLeu][PheTyr]Arg[GluAspVal][ThrAsn] (SEQ ID NO:99)
(n) CysXaaXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:48)
CysXaaXaaAspProXaaCysXaaTrpAsp (SEQ ID NO:49)
CysXaaXaaCysXaaXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:51)
CysXaaXaaCysXaaXaaAspXaaXaaCysXaaTrpAsp (SEQ ID NO:52).

The following 5 classes include peptides which encompass peptides encoded in open reading frames of Variola major or Vaccinia viruses. Accordingly, in the event that these viral peptides are not novel per se, the present invention discloses a hitherto unforseen and unforseeable utility for these peptides as immunosuppressants and targets of anti-viral therapy.

(b)CysGlyThr[AsnGly][AlaSerAsn][TyrPheHisGly][LysArgHisAsnGln] (SEQ ID NO:03)
CysGlyThr[AsnGly][AlaSerAsn]XaaXaaPro (SEQ ID NO:04)
CysGlyThr[AsnGly]XaaXaaXaaProXaa[CysAsp] (SEQ ID NO:05)
CysGlyThrXaaXaaXaaXaaProXaa[CysAsp]XaaXaa[TyrIle] (SEQ ID NO:06)
(f)Asp[LysPheTyr]Val[PheTyr][PheTyrIleLeu][PheTyrIleLeu][PheTyr] (SEQ ID NO:22)
[ValIle][PheTyr][PheTyrIleLeu][PheTyrIleLeu]Phe[ArgThr]Xaa[ThrAsn] (SEQ ID NO:23)
Val[PheTyr][PheTyrIleLeu][PheTyrIleLeu][PheTyr][ArgThr][GluAspVal][ThrAsn] (SEQ ID NO:100)
(i)[ArgLysAsn]Trp[ThrAlaSer][ThrAlaSer][PheTyrLeu]Leu[LysArg] (SEQ ID NO:30)
Trp[ThrAlaSer][ThrAlaSer][PheTyrLeu]LeuLys[AlaSerValIleLeu]XaaLeu (SEQ ID NO:33)

Trp[ThrAlaSer][ThrAlaSer]XaaLeuLysXaaXaaLeuXaaCys (SEQ ID NO:34)

TrpXaa[ThrSer]XaaLeuLysXaaXaaLeuXaaCys (SEQ ID NO:35)

(k) SerAla[ValIleLeu]CysXaa[PheTyr]XaaMet (SEQ ID NO:39)

(m) [ValLeuIle]ProXaaPro[ArgAla]ProGlyXaaCys (SEQ ID NO:42).

The disclosed semaphorin sequence data are used to define a wide variety of other semaphorin- and semaphorin receptor-specific binding agents using immunologic, chromatographic or synthetic methods available to those skilled in the art.

Of particular significance are peptides comprising unique portions of semaphorin-specific receptors and polypeptides comprising a sequence substantially similar to that of a substantially full-length semaphorin receptor. Using semaphorin peptides, these receptors are identified by a variety of techniques known to those skilled in the art where a ligand to the target receptor is known, including expression cloning as set out in the exemplification below. For other examples of receptor isolation with known ligand using expression cloning, see, Staunton et al (1989) Nature 339, 61; Davis et al (1991) Science 253, 59; Lin et al (1992) Cell 68, 775; Gearing et al (1989) EMBO 8, 3667; Aruffo and Seed (1987) PNAS 84, 8573 and references therein. Generally, COS cells are transfected to express a cDNA library or PCR product and cells producing peptides/polypeptides which bind a semaphorin/receptor peptide/polypeptide are isolated. For neurosemaphorin receptors, fetal brain cDNA libraries are preferred; for immunosemaphorin receptors, libraries derived from activated lymphoid or myeloid cell lines or tissue derived from sites of inflammation or delayed-type hypersensitivity are preferred; and for semaphorin and semaphorin receptor variants used by tumor cells to evade immune surveillance or suppress an immune response (oncosemaphorins), libraries derived from cancerous tissue or tumor cell lines resistant to the host immune system are preferred. Alternatively, PCR primers based upon known semaphorin/receptor sequences such as those disclosed herein are used to amplify PCR product from such tissues/cells. Other receptor/ligand isolation methods using immobilized ligand or antibody are known to those skilled in the art.

Semaphorin receptor peptides with receptor binding specificity are identified by a variety of ways including having conserved consensus sequences with other semaphorin receptors, by crosslinking to ligand or receptor-specific antibody, or preferably, by screening such peptides for semaphorin binding or disruption of semaphorin-receptor binding. Methods for identifying semaphorin receptor peptides with the requisite binding activity are described herein or otherwise known to those skilled in the art. By analogous methods, semaphorin receptor peptides are used to define additional semaphorin peptides with semaphorin binding specificity, particularly receptor specificity.

The various semaphorin and semaphorin receptor peptides are used to define

functional domains of semaphorins, identify compounds that associate with semaphorins, design compounds capable of modulating semaphorin-mediated nerve and immune cell function, and define additional semaphorin and semaphorin receptor-specific binding agents. For example, semaphorin mutants, including deletion mutants are generated from the disclosed semaphorin sequences and used to identify regions important for specific protein-ligand or protein-protein interactions, for example, by assaying for the ability to mediate repulsion or preclude aggregation in cell-based assays as described herein. Further, x-ray crystallographic data of the disclosed protein are used to rationally design binding molecules of determined structure or complementarity for modulating growth cone growth and guidance.

Additional semaphorin- and receptor-specific agents include specific antibodies that can be modified to a monovalent form, such as Fab, Fab', or Fv, specifically binding oligopeptides or oligonucleotides and most preferably, small molecular weight organic receptor antagonists. For example, the disclosed semaphorin and receptor peptides are used as immunogens to generate semaphorin- and receptor-specific polyclonal or monoclonal antibodies. See, Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, for general methods. Anti-idiotypic antibody, especially internal imaging anti-ids are also prepared using the disclosures herein.

In addition to semaphorin and semaphorin-receptor derived polypeptides and peptides, other prospective agents are screened from large libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of saccharide, peptide, and nucleic acid based compounds. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily producible. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means. See, e.g. Houghten et al. and Lam et al (1991) Nature 354, 84 and 81, respectively and Blake and Litzi-Davis (1992), Bioconjugate Chem 3, 510.

Useful agents are identified with a range of assays employing a compound comprising the subject peptides or encoding nucleic acids. A wide variety of in vitro, cell-free binding assays, especially assays for specific binding to immobilized compounds comprising semaphorin or semaphorin receptor peptide find convenient use. While less preferred, cell-based assays may be used to determine specific effects of prospective agents on semaphorin-receptor binding may be assayed. Optionally, the intracellular C-terminal domain is substituted with a sequence encoding a oligopeptide or polypeptide domain that provides a detectable intracellular signal upon ligand binding different from the natural receptor. Useful intracellular domains include those of the human insulin receptor and the TCR, especially domains with kinase activity and domains capable of

triggering calcium influx which is conveniently detected by fluorimetry by preloading the host cells with Fura-2. More preferred assays involve simple cell-free in vitro binding of candidate agents to immobilized semaphorin or receptor peptides, or vice versa. See, e.g. Fodor et al (1991) Science 251, 767 for light directed parallel synthesis method. Such assays are amenable to scale-up, high throughput usage suitable for volume drug screening.

Useful agents are typically those that bind to a semaphorin or disrupt the association of a semaphorin with its receptor. Preferred agents are semaphorin-specific and do not cross react with other neural or lymphoid cell membrane proteins. Useful agents may be found within numerous chemical classes, though typically they are organic compounds; preferably small organic compounds. Small organic compounds have a molecular weight of more than 150 yet less than about 4,500, preferably less than about 1500, more preferably, less than about 500. Exemplary classes include peptides, saccharides, steroids, heterocyclics, polycyclics, substituted aromatic compounds, and the like.

Selected agents may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways as described above, e.g.to enhance their proteolytic stability. Other methods of stabilization may include encapsulation, for example, in liposomes, etc.

The subject binding agents may be prepared in a variety of ways known to those skilled in the art. For example, peptides under about 60 amino acids can be readily synthesized today using conventional commercially available automatic synthesizers. Alternatively, DNA sequences may be prepared encoding the desired peptide and inserted into an appropriate expression vector for expression in a prokaryotic or eukaryotic host. A wide variety of expression vectors are available today and may be used in conventional ways for transformation of a competent host for expression and isolation. If desired, the open reading frame encoding the desired peptide may be joined to a signal sequence for secretion, so as to permit isolation from the culture medium. Methods for preparing the desired sequence, inserting the sequence into an expression vector, transforming a competent host, and growing the host in culture for production of the product may be found in U.S. Patent Nos. 4,710,473, 4,711,843 and 4,713,339.

For therapeutic uses, the compositions and agents disclosed herein may be administered by any convenient way, preferably parenterally, conveniently in a pharmaceutically or physiologically acceptable carrier, e.g., phosphate buffered saline, saline, deionized water, or the like. Typically, the compositions are added to a retained physiological fluid such as blood or synovial fluid. For CNS administration, a variety of techniques are available for promoting transfer of the therapeutic across the blood brain barrier including

disruption by surgery or injection, drugs which transciently open adhesion contact between CNS vasculature endothelial cells, and compounds which fascilitate translocation through such cells. As examples, many of the disclosed therapeutics are amenable to directly injected or infused, contained within implants e.g. osmotic pumps, grafts comprising appropriately transformed cells. Generally, the amount administered will be empirically determined, typically in the range of about 10 to 1000 μg/kg of the recipient. For peptide agents, the concentration will generally be in the range of about 50 to 500 μg/ml in the dose administered. Other additives may be included, such as stabilizers, bactericides, etc. These additives will be present in conventional amounts.

The invention provides isolated nucleic acid sequences encoding the disclosed semaphorin and semaphorin receptor peptides and polypeptides, including sequences substantially identical to sequences encoding such polypeptides. An "isolated" nucleic acid sequence is present as other than a naturally occurring chromosome or transcript in its natural state and typically is removed from at least some of the nucleotide sequences with which it is normally associated with on a natural chromosome. A complementary sequence hybridizes to a unique portion of the disclosed semaphorin sequence under low stringency conditions, for example, at 50°C and SSC (0.9 M saline/0.09 M sodium citrate) and that remains bound when subject to washing at 55°C with SSC. Regions of non-identity of complementary nucleic acids are preferably or in the case of homologous nucleic acids, a nucleotide change providing a redundant codon. A partially pure nucleotide sequence constitutes at least about 5%, preferably at least about 30%, and more preferably at least about 90% by weight of total nucleic acid present in a given fraction.

Unique portions of the disclosed nucleic acid sequence are of length sufficient to distinguish previously known nucleic acid sequences. Thus, a unique portion has a nucleotide sequence at least long enough to define a novel oligonucleotide. Preferred nucleic acid portions encode a unique semaphorin peptide. The nucleic acids of the invention and portions thereof, other than those used as PCR primers, are usually at least about 60 bp and usually less than about 60 kb in length. PCR primers are generally between about 15 and 100 nucleotides in length.

Nucleotide (cDNA) sequences encoding several full length semaphorins are disclosed herein. The invention also provides for the disclosed sequences modified by transitions, transversions, deletions, insertions, or other modifications such as alternative splicing and also provides for genomic semaphorin sequences, and gene flanking sequences, including regulatory sequences; included are DNA and RNA sequences, sense and antisense. Preferred DNA sequence portions include portions encoding the preferred amino acid sequence portions disclosed above. For antisense applications where the inhibition of semaphorin expression is indicated, especially useful oligonucleotides are between about 10 and 30

nucleotides in length and include sequences surrounding the disclosed ATG start site, especially the oligonucleotides defined by the disclosed sequence beginning about 5 nucleotides before the start site and ending about 10 nucleotides after the disclosed start site. Other especially useful semaphorin mutants involve deletion or substitution modifications of the disclosed cytoplasmic C-termini of transmembrane semaphorins. Accordingly, semaphorin mutants with semaphorin binding affinities but with altered intracellular signal transduction capacities are produced.

For modified semaphorin-encoding sequences or related sequences encoding proteins with semaphorin-like functions, there will generally be substantial sequence identity between at least a segment thereof and a segment encoding at least a portion of the disclosed semaphorin sequence, preferably at least about 60%, more preferably at least 80%, most preferably at least 90% identity. Homologous segments are particularly within semaphorin domain-encoding regions and regions encoding protein domains involved in protein-protein, particularly semaphorin-receptor interactions and differences within such segments are particularly conservative substitutions.

Typically, the invention's semaphorin peptide encoding polynucleotides are associated with heterologous sequences. Examples of such heterologous sequences include regulatory sequences such as promoters, enhancers, response elements, signal sequences, polyadenylation sequences, etc., introns, 5' and 3' noncoding regions, etc. Other useful heterologous sequences are known to those skilled in the art or otherwise disclosed references cited herein. According to a particular embodiment of the invention, portions of the semaphorin encoding sequence are spliced with heterologous sequences to produce soluble, secreted fusion proteins, using appropriate signal sequences and optionally, a fusion partner such as β-Gal.

The disclosed sequences are also used to identify and isolate other natural semaphorins and analogs. In particular, the disclosed nucleic acid sequences are used as hybridization probes under low-stringency or PCR primers, e.g. oligonucleotides encoding functional semaphorin domains are $^{32}$P-labeled and used to screen λcDNA libraries at low stringency to identify similar cDNAs that encode proteins with related functional domains. Additionally, nucleic acids encoding at least a portion of the disclosed semaphorin are used to characterize tissue specific expression of semaphorin as well as changes of expression over time, particularly during organismal development or cellular differentiation.

The semaphorin encoding nucleic acids can be subject to alternative purification, synthesis, modification, sequencing, expression, transfection, administration or other use by methods disclosed in standard manuals such as Molecular Cloning, A Laboratory Manual (2nd Ed., Sambrook, Fritsch and Maniatis, Cold Spring Harbor), Current Protocols in Molecular Biology (Eds. Aufubel, Brent, Kingston, More, Feidman, Smith and Stuhl, Greene

Publ. Assoc., Wiley-Interscience, NY, NY, 1992) or that are otherwise known in the art. For example, the nucleic acids can be modified to alter stability, solubility, binding affinity and specificity, etc. semaphorin-encoding sequences can be selectively methylated, etc. The nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescers, biotinylation, etc.

The invention also provides vectors comprising nucleic acids encoding semaphorin peptides, polypeptides or analogs. A large number of vectors, including plasmid and viral vectors, have been described for expression in a variety of eukaryotic and prokaryotic hosts. Advantageously, vectors may also include a promotor operably linked to the semaphorin-encoding portion. Vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g. antibiotic resistance. The inserted semaphorin coding sequences may be synthesized, isolated from natural sources, prepared as hybrids, etc. Suitable host cells may be transformed/transfected/infected by any suitable method including electroporation, $CaCl_2$ mediated DNA uptake, viral infection, microinjection, microprojectile, or other methods.

Appropriate host cells include bacteria, archebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. Of particular interest are E. coli, B. subtilis, Saccharomyces cerevisiae, SF9 cells, C129 cells, 293 cells, Neurospora, and CHO, COS, HeLa cells, immortalized mammalian myeloid and lymphoid cell lines, and pluripotent cells, especially mammalian ES cells and zygotes. Preferred replication systems include M13, ColE1, SV40, baculovirus, lambda, adenovirus, AAV, BPV, etc. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, etc. are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced semaphorins or analogs.

For the production of stably transformed cells and transgenic animals, nucleic acids encoding the disclosed semaphorins may be integrated into a host genome by recombination events. For example, such a sequence can be microinjected into a cell, and thereby effect homologous recombination at the site of an endogenous gene, an analog or pseudogene thereof, or a sequence with substantial identity to an semaphorin-encoding gene. Other recombination-based methods such as nonhomologous recombinations, deletion of endogenous gene by homologous recombination, especially in pluripotent cells, etc., provide additional applications. Preferred transgenics and stable transformants over-express the disclosed receptor gene and find use in drug development and as a disease model. Alternatively, knock-out cells and animals find use in development and functional studies.

Methods for making transgenic animals, usually rodents, from ES cells or zygotes are known to those skilled in the art.

The compositions and methods disclosed herein may be used to effect gene therapy. See, e.g. Zhu et al. (1993) Science 261, 209-211; Gutierrez et al. (1992) Lancet 339, 715-721. For example, cells are transfected with semaphorin sequences operably linked to gene regulatory sequences capable of effecting altered semaphorin expression or regulation. To modulate semaphorin translation, cells may be transfected with complementary antisense polynucleotides. For gene therapy involving the transfusion of semaphorin transfected cells, administration will depend on a number of variables that are ascertained empirically. For example, the number of cells will vary depending on the stability of the transfused cells. Transfusion media is typically a buffered saline solution or other pharmacologically acceptable solution. Similarly the amount of other administered compositions, e.g. transfected nucleic acid, protein, etc., will depend on the manner of administration, purpose of the therapy, and the like.

The following examples are offered by way of illustration and not by way of limitation.

## EXAMPLES

I. Isolation and characterization of Grasshopper Semaphorin I (SEQ ID NOS: 57 and 58(previously referred to as Fasciclin IV)

In order to identify cell surface molecules that function in selective fasciculation, a series of monoclonal antibody (MAb) screens was conducted. The immunogen used for most of these screens was membranes from the longitudinal connectives (the collection of longitudinal axons) between adjacent segmental ganglia of the nervous system of the larval grasshopper. From these screens, MAb 3B11 and 8C6 were used to purify and characterize two surface glycoproteins, fasciclin I and fasciclin II, see, Bastiani et al., 1987; the genes encoding both were subsequently cloned, see, Snow et al. 1989, Zinn et al. 1988, and Harrelson and Goodman, 1988.

Another MAb isolated during these screens, MAb 6F8, was chosen for the present study because, just as with fasciclin I and fasciclin II, the antigen recognized by this MAb is expressed on a different but overlapping subset of axon pathways in the developing CNS. The 6F8 antigen appears to be localized on the outside of cell surfaces, as indicated by MAb binding when incubated both in live preparations, and in fixed preparations in which no detergents have been added. Because the 6F8 antigen is a surface glycoprotein expressed on

a subset of axon fascicles (see below), we call it fasciclin IV.

Fasciclin IV expression begins early in embryonic development before axonogenesis. At 29% of development, expression is seen on the surface of the midline mesectodermal cells and around 5-7 neuroblasts and associated ectodermal cells per hemisegment. This expression is reminiscent of the mesectodermal and neuroblast-associated expression observed with both fasciclin I and fasciclin II; however, in each case, the pattern resolves into a different subset of neuroblasts and associated ectodermal cells.

At 32% of development, shortly after the onset of axonogenesis in the CNS, fasciclin IV expression is seen on the surface of the axons and cell bodies of the three pairs of MP4, MP5, and MP6 midline progeny, the three U motoneurons, and on several unidentified neurons in close proximity to the U's. This is in contrast to fasciclin II, which at this stage is expressed on the MP1 and dMP2 neurons, and fasciclin I, which is expressed on the U neurons but not on any midline precursor progeny.

The expression of fasciclin IV on a subset of axon pathways is best observed around 40% of development, after the establishment of the first longitudinal and commissural axon pathways . At this stage, the protein is expressed on two longitudinal axon fascicles, a subset of commissural axon fascicles, a tract extending anteriorly along the midline, and a subset of fascicles in the segmental nerve (SN) and intersegmental nerve (ISN) roots.

Specifically, fasciclin IV is expressed on the U fascicle, a longitudinal pathway (between adjacent segmental neuromeres) pioneered in part by the U neurons, and on the A/P longitudinal fascicle (in part an extension of the U fascicle within each segmental neuromere). In addition, fasciclin IV is also expressed on a second narrower, medial, and more ventral longitudinal pathway. The U axons turn and exit the CNS as they pioneer the ISN; the U's and many other axons within the ISN express fasciclin IV. The continuation of the U fascicle posterior to the ISN junction is also fasciclin IV-positive. The specificity of fasciclin IV for distinct subsets of longitudinal pathways can be seen by comparing fasciclin IV and fasciclin II expression in the same embryo; fasciclin IV is expressed on the U and A/P pathways whereas fasciclin II is expressed on the MP1 pathway.

The axons in the median fiber tract (MFT) also express fasciclin IV. The MFT is pioneered by the three pairs of progeny of the midline precursors MP4, MP5, and MP6. The MFT actually contains three separate fascicles. The axons of the two MP4 progeny pioneer the dorsal MFT fascicle and then bifurcate at the posterior end of the anterior commissure; whereas the axons of the two MP6 progeny pioneer the ventral MFT fascicle and then bifurcate at the anterior end of the posterior commissure. Fasciclin IV is expressed on the cell bodies of the six MP4, MP5, and MP6 neurons, and on their growth cones and axons as they extend anteriorly in the MFT and bifurcate in one of the two commissures. However, this expression is regional in that once these axons bifurcate and begin to extend laterally across

the longitudinal pathways and towards the peripheral nerve roots, their expression of fasciclin IV greatly decreases. Thus, fasciclin IV is a label for the axons in the MFT and their initial bifurcations in both the anterior and posterior commissures. It appears to be expressed on other commissural fascicles as well. However, the commissural expression of fasciclin IV is distinct from the transient expression of fasciclin II along the posterior edge of the posterior commissure, or the expression of fasciclin I on several different commissural axon fascicles in both the anterior and posterior commissure (Bastiani et al., 1987; Harrelson and Goodman, 1988).

Fasciclin IV is also expressed on a subset of motor axons exiting the CNS in the SN. The SN splits into two major branches, one anterior and the other posterior, as it exits the CNS. Two large bundles of motoneuron axons in the anterior branch express fasciclin IV at high levels; one narrow bundle of motoneuron axons in the posterior branch expresses the protein at much lower levels. Fasciclin IV is also expressed on many of the axons in the ISN.

The CNS and nerve root expression patterns of fasciclin IV, fasciclin I, and fasciclin II at around 40% of embryonic development are summarized below. Although there is some overlap in their patterns (e.g., both fasciclin IV and fasciclin I label the U axons), these three surface glycoproteins label distinct subsets of axon pathways in the developing CNS.

**Fasciclin IV is expressed on epithelial bands in the developing limb bud**

Fasciclin IV is expressed on the developing limb bud epithelium in circumferential bands; at 34.5% of development these bands can be localized with respect to constrictions in the epithelium that mark presumptive segment boundaries. In addition to a band just distal to the trochanter/coxa segment boundary, bands are also found in the tibia, femur, coxa, and later in development a fifth band is found in the tarsus. Fasciclin IV is also expressed in the nascent chordotonal organ in the dorsal aspect of the femur. The bands in the tibia, trochanter, and coxa completely encircle the limb. However, the femoral band is incomplete, containing a gap on the anterior epithelia of this segment.

The position of the Ti1 axon pathway with respect to these bands of fasciclin IV-positive epithelia suggests a potential role for fasciclin IV in guiding the Ti1 growth cones. First, the band of fasciclin IV expression in the trochanter, which is approximately three epithelial cell diameters in width when encountered by the Ti1 growth cones, is the axial location where the growth cones reorient from proximal migration to circumferential branch extension. The Tr1 cell, which marks the location of the turn, lies within this band, usually over the central or the proximal cell tier. Secondly, although there is a more distal fasciclin IV expressing band in the femur, where a change in Ti1 growth is not observed, there exists a gap in this band such that fasciclin IV expressing cells are not traversed by the Ti1 growth cones. The Ti1 axons also may encounter a fasciclin IV expressing region within the coxa,

where interactions between the growth cones, the epithelial cells, and the Cx1 guidepost cells have not yet been investigated. In addition to its expression over the surface of bands of epithelial cells, fasciclin IV protein, as visualized with MAb 6F8, is also found on the basal surface of these cells in a punctate pattern. This punctate staining is not an artifact of the HRP immunocytochemistry since fluorescent visualization of MAb 6F8 is also punctate. The non-neuronal expression of fasciclin IV is not restricted to limb buds. Circumferential epithelial bands of fasciclin IV expression are also seen on subesophageal mandibular structures and on the developing antennae.

**MAb directed against fasciclin IV can alter the formation of the Ti1 axon pathway in the limb bud**

The expression of fasciclin IV on an epithelial band at a key choice point in the formation of the Ti1 axon pathway led us to ask whether this protein is involved in growth cone guidance at this location. To answer this question, we cultured embryos, or epithelial fillets (e. g., O'Connor et al., 1990), during the 5% of development necessary for normal pathway formation, either in the presence or absence of MAb 6F8 or 6F8 Fab fragments. Under the culture conditions used for these experiments, defective Ti1 pathways are observed in 14% of limbs (Chang et al., 1992); this defines the baseline of abnormalities observed using these conditions. For controls we used other MAbs and their Fab fragments that either bind to the surfaces of these neurons and epithelial cells (MAb 3B11 against the surface protein fasciclin I) or do not (MAb 4D9 against the nuclear protein engrailed; Patel et al., 1989). To assess the impact of MAb 6F8 on Ti1 pathway formation, we compared the percentage of aberrant pathways observed following treatment with MAb 6F8 to that observed with MAbs 3B11 and 4D9. Our cultures began at 32% of development when the Ti1 growth cones have not yet reached the epithelium just distal to the trochanter/coxa boundary and therefore have not encountered epithelial cells expressing fasciclin IV. Following approximately 30 hours in culture (~4% of development), embryos were fixed and immunostained with antibodies to HRP in order to visualize the Ti1 axons and other neurons in the limb bud. Criteria for scoring the Ti1 pathway, and the definition of "aberrant", are described in detail in the Experimental Procedures.

Although MAb 6F8 does not arrest pathway formation, several types of distinctive, abnormal pathways are observed. These defects generally begin where growth cones first contact the fasciclin IV expressing cells in the trochanter. Normally, the Ti1 neurons each have a single axon, and the axons of the two cells are fasciculated in that portion of the pathway within the trochanter. Following treatment with MAb 6F8, multiple long axon branches are observed within, and proximal to, the trochanter. Two major classes of pathways are taken by these branches; in 36% of aberrant limbs, multiple, long axon branches

extend ventrally in the region distal to the Cx1 cells which contains the band of fasciclin IV expressing epithelial cells. In the ventral region of the trochanter, these branches often independently turn proximally to contact the Cx1 cells, and thus complete the pathway in this region.

In the second major class of pathway defect, seen in 47% of aberrant limbs, axon branches leave the trochanter at abnormal, dorsal locations, and extend proximally across the trochanter/coxa boundary. These axons then veer ventrally, often contacting the Cx1 neurons. The remaining 17% of defects include defasciculation distal to the trochanter, axon branches that fail to turn proximally in the ventral trochanter and continue into the posterior compartment of the limb, and axon branches which cross the trochanter/coxa boundary and continue to extend proximally without a ventral turn.

When cultured in the presence of MAb 6F8, 43% of limbs exhibited malformed Ti1 pathways (n = 381) as compared to 11% with MAb 3B11 (n = 230) and 5% with MAb 4D9 (n = 20). These percentages are pooled from treatments with MAbs concentrated from hybridoma supernatant, IgGs isolated from these supernatants, and Fab fragments isolated from these IgG preparations (see Experimental Procedures). The frequency of malformed Ti1 pathways and the types of defects observed showed no significant variation regardless of the method of antibody preparation or type of antibody used. Since Fabs show similar results as IgGs, the effects of MAb 6F8 are not due to cross linking by the bivalent IgG.

In summary, following treatment with MAb 6F8, the Ti1 pathway typically exhibits abnormal morphology beginning just distal to the trochanter and at the site of fasciclin IV expression. The two most common types of Ti1 pathway defects described above occur in 36% of experimental limbs (treated with MAb 6F8), but are seen in only 4% of control limbs (treated with MAbs 3 B11 and 4D9).

**Fasciclin IV cDNAs encode a novel integral membrane protein**

Grasshopper fasciclin IV was purified by passing crude embryonic grasshopper lysates over a MAb 6F8 column. After affinity purification, the protein was eluted, precipitated, denatured, modified at cysteines, and digested with either trypsin or Lys-C. Individual peptides were resolved by reverse phase HPLC and microsequenced using standard methods.

The amino acid sequences derived from these proteolytic fragments were used to generate oligonucleotide probes for PCR experiments, resulting in products that were used to isolate cDNA clones from the Zinn embryonic grasshopper cDNA library (Snow et al., 1988). Sequence analysis of these cDNAs reveals a single open reading frame (ORF) encoding a protein with two potential hydrophobic stretches of amino acids: an amino-terminal signal sequence of 20 residues and (beginning at amino acid 627) a potential transmembrane domain of 25 amino acids. Thus, the deduced protein has an extracellular domain of 605 amino

acids, a transmembrane domain, and a cytoplasmic domain of 78 amino acids. The calculated molecular mass of the mature fasciclin IV protein is 80 kd and is confirmed by Western blot analysis of the affinity purified and endogenous protein as described below. The extracellular domain of the protein includes 16 cysteine residues that fall into three loose clusters but do not constitute a repeated domain and are not similar to other known motifs with cysteine repeats. There are also six potential sites for N-linked glycosylation in the extracellular domain. Treatment of affinity purified fasciclin IV with N-Glycanase demonstrates that fasciclin IV does indeed contain N-linked oligosaccharides. Fasciclin IV shows no sequence similarity when compared with other proteins in the PIR data base using BLASTP (Altschul et al., 1990), and is therefore a novel type I integral membrane protein.

A polyclonal antiserum directed against the cytoplasmic domain of the protein encoded by the fasciclin IV cDNA was used to stain grasshopper embryos at 40% of development. The observed staining pattern was identical to that seen with MAb 6F8. On Western blots, this antiserum recognizes the protein we affinity purified using MAb 6F8 and then subjected to microsequence analysis. Additionally, the polyclonal serum recognizes a protein of similar molecular mass from grasshopper embryonic membranes. Taken together these data indicate that the sequence we have obtained is indeed fasciclin IV.

Four other cell surface proteins that label subsets of axon pathways in the insect nervous system (fasciclin I, fasciclin II, fasciclin III, and neuroglian) are capable of mediating homophilic cell adhesion when transfected into S2 cells in vitro (Snow et al., 1989; Elkins et al., 1990b; Grenningloh et al., 1990). To ask whether fasciclin IV can function as a homophilic cell adhesion molecule, the fasciclin IV cDNA with the complete ORF was placed under the control of the inducible metallothionein promoter (Bunch et al., 1988), transfected into S2 cells, and assayed for its ability to promote adhesion in normally non-adhesive S2 cells. Following induction with copper, fasciclin IV was synthesized in these S2 cells as shown by Western blot analysis and cell surface staining of induced S2 cells with the polyclonal antiserum described above.

We observed no evidence for aggregation upon induction of fasciclin IV expression, thus suggesting that, in contrast to the other four proteins, fasciclin IV does not function as a homophilic cell adhesion molecule. Alternatively, fasciclin IV-mediated aggregation might require some further posttranslational modification, or co-factor, not supplied by the S2 cells, but clearly this protein acts differently in the S2 cell assay than the other four axonal glycoproteins previously tested. This is consistent with the pattern of fasciclin IV expression in the embryonic limb since only the epithelial cells and not the Ti1 growth cones express fasciclin IV, and yet antibody blocking experiments indicate that fasciclin IV functions in the epithelial guidance of these growth cones. Such results suggest that fasciclin IV functions in a heterophilic adhesion or signaling system.

**Discussion**

Fasciclin IV is expressed on groups of axons that fasciculate in the CNS, suggesting that, much like other insect axonal glycoproteins, it functions as a homophilic cell adhesion molecule binding these axons together. Yet, in the limb bud, fasciclin IV is expressed on a band of epithelium but not on the growth cones that reorient along this band, suggesting a heterophilic function. That fasciclin IV functions in a heterophilic rather than homophilic fashion is supported by the lack of homophilic adhesion in S2 cell aggregation assays. In contrast, fasciclin I, fasciclin II, fasciclin III, and neuroglian all can function as homophilic cell adhesion molecules (Snow et al., 1989; Elkins et al., 1990b; Grenningloh et al., 1990).

cDNA sequence analysis indicates that fasciclin IV is an integral membrane protein with a novel sequence not related to any protein in the present data base. Thus, fasciclin IV represents a new type of protein that functions in the epithelial guidance of pioneer growth cones in the developing limb bud. Given its expression on a subset of axon pathways in the developing CNS, fasciclin IV functions in the guidance of CNS growth cones as well.

The results from the MAb blocking experiments illuminate several issues in Ti1 growth cone guidance and axon morphogenesis in the limb. First, the most striking change in growth cone behavior in the limb is the cessation of proximal growth and initiation of circumferential extension of processes upon encountering the trochanter/coxa boundary region (Bentley and Caudy, 1983; Caudy and Bentley, 1987). This could be because the band of epithelial cells within the trochanter promotes circumferential growth, or because the cells comprising the trochanter/coxa boundary and the region just proximal to it are non-permissive or aversive for growth cone migration, or both. The extension of many axon branches across the trochanter/coxa boundary following treatment with MAb 6F8 suggests that the trochanter/coxa boundary cells, which do not express fasciclin IV, are not aversive or non-permissive. Thus the change in behavior at the boundary appears to be due to the ability of fasciclin IV expressing epithelial cells to promote circumferential extension of processes from the Ti1 growth cones.

Secondly, treatment with MAb 6F8 results in frequent defasciculation of the axons of the two Ti1 neurons, and also formation of abnormal multiple axon branches, within the trochanter over fasciclin IV-expressing epithelial cells. Previous studies have shown that treatment with antibodies against ligands expressed on non-neural substrates (Landmesser et al., 1988), or putative competitive inhibitors of substrate ligands (Wang and Denburg, 1992) can promote defasciculation and increased axonal branching. Our results suggest that Ti1 axon:axon fasciculation and axon branching also are strongly influenced by interactions with substrate ligands, and that fasciclin IV appears to be a component of this interaction within the trochanter.

Thirdly, despite the effects of MAb 6F8 on axon branching, and on crossing the

trochanter/coxa boundary, there remains a pronounced tendency for branches to grow ventrally both within the trochanter and within the distal region of the coxa. Consequently, all signals which can promote ventral migration of the growth cones have not been blocked by MAb 6F8 treatment. Antibody treatment may have a threshold effect in which ventral growth directing properties of fasciclin IV are more robust, and less incapacitated by treatment, than other features; alternatively, guidance information promoting ventral migration may be independent of fasciclin IV. Time lapse video experiments to determine how the abnormal pathways we observe actually form can resolve these issues.

These results demonstrate that fasciclin IV functions as a guidance cue for the Ti1 growth cones just distal to the trochanter/coxa boundary, is required for these growth cones to stop proximal growth and spread circumferentially, and that the function of fasciclin IV in Ti1 pathway formation result from interactions between a receptor/ligand on the Ti1 growth cones and fasciclin IV on the surface of the band of epithelial cells results in changes in growth cone morphology and subsequent reorientation. Fasciclin IV appears to elicit this change in growth cone morphology and orientation via regulation of adhesion, a signal transduction function, or a combination of the two.

## Experimental Procedures

### Immunocytochemistry

Grasshopper embryos were obtained from a colony maintained at the U.C. Berkeley and staged by percentage of total embryonic development (Bentley et al., 1979). Embryos were dissected in PBS, fixed for 40 min in PEM-FA [0.1 M PIPES (pH6.95), 2.0 mM EGTA, 1.0 mM $MgSO_4$, 3.7% formaldehyde], washed for 1 hr with three changes in PBT (1x PBS, 0.5% Triton X-100, 0.2% BSA), blocked for 30 min in PBT with 5% normal goat serum, and incubated overnight at 4°C in primary antibody. PBSap (1x PBS, 0.1% Saponin, o.2% BSA) was used in place of PBT with MAb 8G7. Antibody dilutions were as follows: MAb 6F8 1:1, polyclonal antisera directed against a fasciclin IV bacterial fusion protein (#98-3) 1:400; MAb 8G7 1:4; MAb 8C6 1:1. The embryos were washed for one hour in PBT with three changes, blocked for 30 min, and incubated in secondary antibody for at least 2 hr at room temperature. The secondary antibodies were HRP-conjugated goat anti-mouse and anti-rat IgG (Jackson Immunoresearch Lab), and were diluted 1:300 . Embryos were washed in PBT for one hour with three changes and then reacted in 0.5% diaminobenzidine (DAB) in PBT. The reaction was stopped with several washes in PBS and the embryos were cleared in a glycerol series (50%, 70%, 90%), mounted and viewed under Nomarski or bright field optics. For double-labelled preparations the first HRP reaction was done in PBT containing 0.06% NiCl, followed by washing, blocking, and incubation overnight in the second primary antibody.

The second antibody was visualized with a DAB reaction as described above. Embryos cultured in the presence of monoclonal antibodies were fixed and incubated overnight in goat anti-HRP (Jackson Immunoresearch Labs) conjugated to RITC (Molecular Probes), washed for one hour in PBT with three changes, mounted in 90% glycerol, 2.5% DABCO (Polysciences), and viewed under epifluorescence. S2 cells were stained with polyclonal sera #98-3 diluted 1:400 and processed as described previously (Snow et al., 1989).

**Monoclonal Antibody Blocking Experiments**

In order to test for functional blocking, monoclonal antibody reagents were prepared as follows. Hybridoma supernatant was brought to 20% with $H_2O$-saturated $(NH_4)_2SO_4$, incubated in ice 1 hr, and spun at 15,000 *g* at 4°C for 20 min. The supernatant was brought to 56% with $H_2O$-saturated $(NH_4)_2SO_4$, incubated overnight at 4°C, spun as above. The pellet was resuspended in PBS using approximately 1/40 volume of the original hybridoma supernatant (often remaining a slurry) and dialyzed against 1x PBS overnight at 4°C with two changes. This reagent is referred to as "concentrated hybridoma supernatant." Purified IgG was obtained by using Immunopure Plus Immobilized Protein A IgG Purification Kit (Pierce) to isolate IgG from the concentrated hybridoma supernatant. Fab fragments were obtained using the ImmunoPure Fab Preparation Kit (Pierce) from the previously isolated IgGs. For blocking experiments each reagent was diluted into freshly made supplemented RPMI culture media (O'Connor et al., 1990) and dialyzed overnight at 4°C against 10 volumes of the same culture media. Dilutions were as follows: concentrated hybridoma supernatant 1:4; purified IgG 150mg/ml; Fab 75mg/ml.

Embryos for culture experiments were carefully staged to between 31 and 32% of development. As embryos in each clutch typically differ by less that 1% of embryonic development from each other, the growth cones of the Ti1 neurons at the beginning of the culture period were located approximately in the mid-femur, well distal to the trochanter/coxa segment boundary. From each clutch at least two limbs were filleted and the Ti1 neurons labelled with the lipophillic dye Di I (Molecular Probes) as described (O'Connor et al., 1990) in order to confirm the precise location of the Ti1 growth cones. Prior to culturing, embryos were sterilized and dissected (Chang et al., 1992). The entire amnion and dorsal membrane was removed from the embryo to insure access of the reagents during culturing. Embryos were randomly divided into groups and cultured in one of the blocking reagents described above. Cultures were incubated with occasional agitation at 30°C for 30 hrs. At the end of the culture period embryos were fixed and processed for analysis as described above in immunocytochemistry.

For each culture experiment, the scoring of the Ti1 pathway in each limb was confirmed independently by a second observer. There was no statistically significant

variation between the two observers. Limbs from MAb cultured embryos were compared to representative normal limbs from non-MAb cultured embryos and were scored as abnormal if any major deviation from the normal Ti1 pathway was observed. The Ti1 pathway was scored as abnormal for one or more of the following observed characteristics: (1) defasciculation for a minimum distance of approximately 25 mm anywhere along the pathway, (2) multiple axon branches that extended ventrally within the trochanter, (3) presence of one or more axon branches that crossed the trochanter/coxa boundary dorsal to the Cx1 cells, but then turned ventrally in the coxa and contacted the Cx1 cells, (4) the presence of axon branches that crossed the trochanter/coxa segment boundary, did not turn ventrally, but continued proximally toward the CNS, and (5) failure of ventrally extended axons within the trochanter to contact and reorient proximally to the Cx1 cells. For each MAb tested, the data are presented as a percentage of the abnormal Ti1 pathways observed.

**Protein Affinity Purification and Microsequencing**

Grasshopper fasciclin IV was purified by passing crude embryonic grasshopper lysate (Bastiani et al., 1987) over an Affi-Gel 15 column (Bio Rad) conjugated with the monoclonal antibody 6F8. Protein was eluted with 50 mM DEA (pH 11.5), 0.1% Lauryldimethylamine oxide (Cal Bio Chem), and 1mM EDTA. Protein was then precipitated, denatured, modified at cysteines, and digested with either trypsin or Lys-C (Boehringer-Mannheim). Individual peptides were resolved by RP-HPLC and microsequenced (Applied Biosystems 4771 Microsequencer) using standard chemistry.

**PCR Methods**

DNA complementary to poly(A)+ RNA from 45%-50% grasshopper embryos was prepared (Sambrook et al., 1989). PCR was performed using Perkin Elmer Taq polymerase (Saiki et al., 1988), and partially degenerate (based on grasshopper codon bias) oligonucleotides in both orientations corresponding to a portion of the protein sequence of several fasciclin IV peptides as determined by microsequencing. These oligonucleotides were designed so as not to include all of the peptide-derived DNA sequence, leaving a remaining 9-12 base pairs that could be used to confirm the correct identity of amplified products. All possible combinations of these sequences were tried. 40 cycles were performed, the parameters of each cycle as follows: 96° for one min; a sequentially decreasing annealing temperature (2°C/cycle, starting at 65°C and ending at 55°C for remaining 35 cycles) for 1 min; and at 72°C for one min. Reaction products were cloned into the Sma site of M13 mp10 and sequenced. Two products, 1074 bp and 288 bp in length, contained DNA 3' to the oligonucleotide sequences encoded the additional amino acid sequence of the fasciclin IV peptide from which the oligonuceotides were derived.

**cDNA Isolation and Sequence Analysis**

Both PCR products were used to screen 1 X $10^6$ clones from a grasshopper embryonic cDNA library (Snow et al., 1988). 21 clones that hybridized to both fragments were recovered, and one 2600 bp clone was sequenced using the dideoxy chain termination method (Sanger et al., 1977) and Sequenase (US Biochemical Corp.). Templates were made from M13 mp10 vectors containing inserts generated by sonication of plasmid clones. One cDNA was completely sequenced on both strands using Oligonucleotides and double strand sequencing of plasmid DNA (Sambrook et al., 1989) to fill gaps. Two additional cDNAs were analyzed by double strand sequencing to obtain the 3' 402 bp of the transcript. All three cDNAs were used to construct a plasmid containing the entire transcript. The complete transcript sequence is 2860 bp in length with 452 bp of 5' and 217 bp of 3' untranslated sequences containing stop codons in all reading frames. The predicted protein sequence was analyzed using the FASTDB and BLASTP programs (Intelligenetics). The fasciclin IV ORF unambiguously contains 10 of the 11 peptide sequences determined by microsequencing the fasciclin IV trypsin and Lys-C peptides.

**Generation of Polyclonal Antibodies From Bacterial Fusion Proteins**

Bacterial trpE fusion proteins were constructed using pATH (Koerner et al., 1991) vectors, three restriction fragments encoding extracellular sequences, and one fragment (770 bp HindIII/ Eco R1, which includes amino acids 476-730 ) encoding both extracellular and intracellular sequences (designated #98-3). Fusion proteins were isolated by making an extract of purified inclusion bodies (Spindler et al., 1984), and rats were immunized with ~70mg of protein emulsified in RIBI adjuvant (Immunochem Research). Rats were injected at two week intervals and serum was collected 7 days following each injection. Sera were tested histologically on grasshopper embryos at 45% of development. Construct #98-3 showed a strong response and exhibited a staining pattern identical to that of MAb 6F8. Two of the extracellular constructs responded weakly but also showed the fasciclin IV staining pattern. All pre-immune sera failed to stain grasshopper embryos.

**S2 Cell Transfections, Aggregation Assays, and Western Analysis**

A restriction fragment containing the full length fasciclin IV cDNA was cloned into pRmHa-3 (Bunch et al, 1988) and co-transformed into Drosophila S2 cells (Schneider, 1972) with the plasmid pPC4 (Jokerst et al., 1989), which confers a-amanitin resistance. S2 cells were transformed using the Lipofectin Reagent and recommended protocol (BRL) with minor modifications. All other S2 cell manipulations are essentially as described (Snow et al.,1989), including adhesion assays. Fasciclin IV expression in transformed cell lines was induced for adhesion assays and histology by adding $CuSO_4$ to 0.7 mM and incubating for at

least 48 hrs. Northern analysis confirmed transcription of fasciclin IV and surface-associated staining of the S2 cells with polyclonal serum #98-3 strongly suggests fasciclin IV is being transported to the cell surface. Preparation of membranes from S2 cells and from grasshopper embryos, PAGE, and Western blot were performed as previously described (Elkins et al., 1990b) except that signal was detected using the enhanced chemiluminescence immunodetection system kit (Amersham). Amount of protein per lane in each sample loaded: fasciclin IV protein, ~5 ng; S2 cell membranes, 40 mg; grasshopper membranes 80 mg. Amounts of protein loaded were verified by Ponceau S staining of the blot prior to incubation with the antibody.

**References cited in Example I**


Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Bastiani, M. J., de Couet, H. G., Quinn, J. M. A., Karlstrom, R. O., Kotrla, K., Goodman, C. S., and Ball, E. E. (1992). Position-specific expression of the annulin protein during grasshopper embryogenesis. Dev. Biol., in press.

Bastiani, M.J., du Lac, S., and Goodman, C. S. (1986). Guidance of neuronal growth cones in the grasshopper embryo. I. Recognition of a specific axonal pathway by the pCC neuron. J. Neurosci. *6*, 3518-3531.

Bastiani, M. J., and Goodman, C. S. (1986). Guidance of neuronal growth cones in the grasshopper embryo: III. Recognition of specific glial pathways. J. Neurosci. *6*, 3542-3551.

Bastiani, M. J., Harrelson, A. L., Snow, P. M., and Goodman, C. S. (1987). Expression of fasciclin I and II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. Cell *48*, 745-755.

Bastiani, M. J., Raper, J. A., and Goodman, C. S. (1984). Pathfinding by neuronal growth cones in grasshopper embryos. III. Selective affinity of the G growth cone for the P cells within the A/P fascicle. J. Neurosci. *4*, 2311-2328.

Bentley, D., and Caudy, M. (1983). Pioneer axons lose directed growth after selective killing of guidepost cells. Nature. *304*, 62-65.

Bentley, D., Keshishian, H., Shankland, M., and Toroian-Raymond, A. (1979). Quantitative staging of embryonic development of the grasshopper, Schistocerca nitens. J. Embryol. Exp. Morph. *54*, 47-74.

Bentley, D., and O'Connor, T.P. (1992). Guidance and steering of peripheral pioneer growth cones in grasshopper embryos. In The Nerve Growth Cone, P.C.Letourneau, S. B.Kater, and E.R. Macagno eds. (New York: Raven Press, Ltd.), pp. 265-282.

Bunch, T. A., Grinblat, Y., and Goldstein, L.S.B. (1988). Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells. Nucleic Acids Res. *16*, 1043-1061.

Chang, W.S., Serikawa, K., Allen, K., and Bentley, D. (1992). Disruption of pioneer growth cone guidance in vivo by removal of glycosyl-phosphatidylinositol-anchored cell surface proteins. Development. *114*, 507-519.

Caudy, M., and Bentley, D. (1987). Pioneer growth cone behavior at a differentiating limb segment boundary in the grasshopper embryo. Dev. Biol. *119*, 454-465.

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. Biochemistry. *13*, 222-245.

Elkins, T., Zinn, K., McAllister, L., Hoffmann, F. M., and Goodman, C. S. (1990a). Genetic analysis of a Drosophila neural cell adhesion molecule: Interaction of fasciclin I and abelson tyrosine kinase mutations. Cell. *60*, 565-575.

Elkins, T., Hortsch, M., Bieber, A. J., Snow, P. M., and Goodman, C. S. (1990b). Drosophila fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. J. Cell Biol. *110*, 1825-1832.

Goodman, C. S., Bate, C. M., and Spitzer, N. C. (1981). Embryonic development of identified neurons: Origin and transformation of the H cell. J. Neurosci. *1*, 94-102.

Grenningloh, G., Bieber, A., Rehm, J., Snow, P. M., Traquina, Z., Hortsch, M., Patel, N. H., and Goodman, C. S. (1990). Molecular genetics of neuronal recognition in Drosophila: Evolution and function of immunoglobulin superfamily cell adhesion molecules. Cold Spring Harbor Symp. Quant. Biol. *55*, 327-340.

Grenningloh, G., Rehm, E. J., and Goodman, C. S. (1991). Genetic analysis of growth cone guidance in Drosophila: Fasciclin II functions as a neuronal recognition molecule. Cell. *67*, 45-57.

Harrelson, A. L., and Goodman, C. S. (1988). Growth cone guidance in insects: Fasciclin II is a member of the immunoglobulin superfamily. Science. *242*, 700-708.

Jacobs, J. R., and Goodman, C. S. (1989). Embryonic development of axon pathways in the Drosophila CNS. I. A glial scaffold appears before the first growth cones. J. Neurosci. *7*, 2402-2411.

Jay, D. J., and Keshishian, H. (1990). Laser inactivation of fasciclin I disrupts axon adhesion of grasshopper pioneer neurons. Nature. *348*, 548-551.

Jokerst, R. S., Weeks, J. R. Zehring, W. A., and Greenleaf, A. L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. Mol. Gen. Genet. *215*, 266-275.

Koerner, T. J., Hill, J. E., Myers, A. M., and Tzagoloff, A. (1991). High-expression vectors with multiple cloning sites for construction of trpE-fusion genes: pATH vectors. Methods Enzymol. *194*, 477-490.

Landmesser, L., Dahm L., Schultz, K., and Rutishauser, U. (1988). Distinct roles for adhesion molecules during innervation of embryonic chick muscle. Dev. Biol. *130*, 645-670.

Lefcort, F., and Bentley, D. (1987). Pathfinding by pioneer neurons in isolated, opened and mesoderm-free limb buds of embryonic grasshoppers. Dev. Biol. *119*, 466-480.

Lefcort, F., and Bentley, D. (1989). Organization of cytoskeletal elements and organelles preceding growth cone emergence from an identified neuron in situ. J. Cell. Biol. *108*, 1737-1749.

O'Connor, T. P., Duerr, J. S., and Bentley, D. (1990). Pioneer growth cone steering decisions mediated by single filopodial contacts *in situ* . J. Neurosci. *10*, 3935-3946.

Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B., and Goodman, C. S. (1989). Expression of *engrailed* proteins in arthropods, annelids, and

chordates. Cell. *58*, 955-968.

Patel, N. H., Snow, P. M., and Goodman, C. S. (1987). Characterization and cloning of fasciclin III: A glycoprotein expressed on a subset of neurons and axon pathways in Drosophila. Cell. *48*, 975-988.

Raper, J. A., Bastiani, M. J., and Goodman, C. S. (1984). Pathfinding by neuronal growth cones in grasshopper embryos. IV. The effects of ablating the A and P axons upon the behavior of the G growth cone. J. Neurosci. *4*, 2329-2345.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487-494.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. *74*, 5463-5467.

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. *27*, 353-365.

Snow, P. M., Bieber, A. J., and Goodman, C. S. (1989). Fasciclin III: a novel homophilic adhesion molecule in Drosophila. Cell. *59*, 313-323.

Snow, P. M., Zinn, K., Harrelson, A. L., McAllister, L., Schilling, J., Bastiani, M. J., Makk, G. , and Goodman, C. S. (1988). Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. Proc. Natl. Acad. Sci. USA *85*, 5291-5295.

Spindler, K. R., Rosser, D. S., and Berk, A. J. (1984). Analysis of adenovirus transforming proteins from early regions 1A and 1B with antisera to inducible fusion antigens produced in Escherichia coli. J. Virol. *49*, 132-141.

Wang, L., and Denburg, J. L. (1992). A role for proteoglycans in the guidance of a subset of pioneer axons in cultured embryos of the cockroach. (1992). Neuron. *8*, 701-714.

Wang, L. S., Feng, Y., and Denburg, J. L. (1992). A multifunctional cell surface developmental stage-specific antigen in the cockroach embryo: involvement in pathfinding by CNS pioneer axons. J. Cell Biol. *118*, 163-176.

Zinn, K., McAllister, L., and Goodman, C. S. (1988). Sequence analysis and neuronal expression of fasciclin I in grasshopper and Drosophila. Cell. *53*, 577-587.

Genbank Accession Number:
The accession number for the sequence reported in this paper is L00709.

II. Isolation and characterization of Tribolium (SEQ ID NOS:63 and 64) and Drosophila (SEQ ID NOS:59 and 60) Semaphorin I, Drosophila Semaphorin II (SEQ ID NOS:61 and 62), Human Semaphorin III (SEQ ID NOS:53 and 54) and Vaccinia Virus Semaphorin IV (SEQ ID NOS:55 and 56) and Variola Major (smallpox) Virus Semaphorin V (SEQ ID NOS:67 and 68).

We used our G-Semaphorin I cDNA in standard low stringency screening methods (of both cDNA and genomic libraries) in an attempt to isolate a potential Semaphorin I homologue from *Drosophila*. We were unsuccessful in these screens. Since the sequence was novel and shared no similarity to anything else in the data base, we then attempted to see if we could identify a Semaphorin I homologue in other, more closely related insects. If possible, we would then compare these sequences to find the most conserved regions, and then to use probes (i.e., oligonucleotide primers for PCR) based on these conserved regions to find a *Drosophila* homologue.

In the process, we used the G-Semaphorin I cDNA in low stringency screens to clone Semaphorin I cDNAs from libraries made from locust *Locusta migratoria* embryonic RNA and from a cDNA embryonic library from the cricket *Acheta domestica*. We used PCR to clone genomic fragments from genomic DNA in the beetle *Tribolium*, and from the moth *Manduca*. We then used the *Tribolium* genomic DNA fragment to isolate cDNA clones and ultimately sequenced the complete ORF for the *Tribolium* cDNA.

In the meantime, we used the partial *Tribolium* and *Manduca* sequences in combination with the complete grasshopper sequence to identify conserved regions that allowed us to design primers for PCR in an attempt to clone a *Drosophila* Semaphorin I homologue. Several pairs of primers generated several different bands, which were subcloned and sequenced and several of the bands gave partial sequences of the Drosophila Semaphorin homologue. One of the bands gave a partial sequence of what was clearly a different, more divergent gene, which we call D-Semaphorin II.

Based on the sequence of PCR products, we knew we had identified two different Drosophila genes, one of which appeared to be the Semaphorin I homologue, and the other a second related gene. The complete ORF sequence of the D-Semaphorin I homologue revealed an overall structure identical to G-Semaphorin I: a signal sequence, an extracellular domain of around 550 amino acids containing 16 cysteines, a transmembrane domain of 25 amino acids, and a cytoplasmic domain of 117 amino acids. When we had finished the sequence for D-Semaphorin II, we were able to begin to run homology searches in the data base, which revealed some of its structural features further described herein. The Semaphorin II sequence revealed a different structure: a signal sequence of 16 amino acids, a ~525 amino acid domain containing 16 cysteines, with a single immunoglobulin (Ig) domain of 66 amino acids, followed by a short unique region of 73 amino acids. There is no evidence for either a transmembrane domain or a potential phospholipid linkage in the C-terminus of this protein. Thus, it appears that the D-Semaphorin II protein is secreted from the cells that produce it. The grasshopper, *Tribolium*, and *Drosophila* Semaphorin I cDNA sequences, as well as the sequence of the D-Semaphorin II cDNA, are shown herein. In addition, we used this same technique to identify Semaphorin I genes in a moth, Manduca sexta, a locust, Locusta migratoria, and a cricket, Acheta domestica.

With this large family of insect Semaphorin genes, we identified a number of good stretches of the right amino acids (with the least degeneracy based on their codons) with strong homology for designing primers for PCR to look for human genes. We designed a set of oligonucleotide primers, and plated out several human cDNA libraries: a fetal brain library (Stratagene), and an adult hippocampus library. We ultimately obtained a human cDNA PCR bands of the right size that did not autoprime and thus were good candidates to be bonafide Semaphorin-like cDNAs from humans. These bands were purified, subcloned, and sequenced.

Whole-mount in situ hybridization experiments showed that D-Semaphorin I and II are expressed by different subsets of neurons in the embryonic CNS. D-Semaphorin I is expressed by certain cells along the midline as well as by other neurons, whereas D-Semaphorin II is not expressed at the midline, but is expressed by a different subset of neurons. In addition, D-Semaphorin II is expressed by a subset of muscles prior to and during the period of innervation by specific motoneuron. On the polytene chromosomes, the D-Semaphorin I gene maps to (gene-band-chromosome) 29E1-22L and that of D-Semaphorin II to 53C9-102R. We have identified loss of function mutations in the D-Semaphorin I gene and a pair of P-element transposon insertions in the D-Semaphorin II gene which appear to cause severe phenotypes.

When we lined up the G-Semaphorin I, T-Semaphorin I, D-Semaphorin I, and D-Semaphorin II sequences and ran the sequences through a sequence data base in search of

other sequences with significant similarity, we discovered a curious finding: these Semaphorins share sequence similarity with the A39R open reading frame (ORF) from Vaccinia virus and the A43R ORF from Variola Major (smallpox) virus and we discovered that the amino acids shared with the virus ORF were in the same regions where the insect proteins shared their greatest similarity. The viral ORF began with a putative signal sequence, continued for several hundred amino acids with sequence similarity to the Semaphorin genes, and then ended without any membrane linkage signal (suggesting that the protein as made by the infected cell would likely be secreted).

We reasoned that the virus semaphorins were appropriated host proteins advantageously exploited by the viruses, which would have host counterparts that most likely function in the immune system to inhibit or decrease an immune response, just as in the nervous system they appear to function by inhibiting growth cone extension. Analogous to situations where viruses are thought to encode a secreted form of a host cellular receptor, here the virus may cause the infected cell to make a lot of the secreted ligand to mimic an inhibitory signal and thus help decrease the immune response.

III. Isolation and characterization of Murine CNS Semaphorin III Receptor using Epitope Tagged Human Semaphorin III (hSIII)

mRNA was isolated from murine fetal brain tissue and used to construct a cDNA library in a mammalian exprssion vector, pCMX, essentially as in Davis et al. (1991) Science 253, 59.

The transfection and screening procedure is modified from Lin et al (1992) Cell 68, 775. COS cells grown on glass slide flaskettes are transfected with pools of the cDNA clones, allowed to bind radioiodinated hSIII truncated at the C-terminus end of the semaphorin domain. In parallel, similarly treated COS cells are allowed to bind unlabelled human semaphorin III truncated at the C-terminus end of the semaphorin domain and there joined to a 10-amino acid extension derived from the human c-myc proto-oncogene product. This modified hSIII allows the identification of hSIII receptors with the use of the tagged ligand as a bridge between the receptor and a murine monoclonal antibody which is specific for an epitope in the c-myc tag. Accordingly, after binding unlabelled hSIII the cells are exposed to the monoclonal which may be labeled directly or subsequently decorated with a secondary anti-mouse labeled antibody for enhanced signal amplification.

Cells are then fixed and screened using dark-field microscopy essentially as in Lin et al. (supra). Positive clones are identified and sequence analysis of murine CNS Semphorin III receptor cDNA clones by the dideoxy chain termination method is used to construct full-length receptor coding sequences.

It is evident from the above results that one can use the methods and compositions

disclosed herein for making and identifying diagnostic probes and therapeutic drugs. It will also be clear to one skilled in the art from a reading of this disclosure that advantage can be taken to effect alterations of semaphorin responsiveness in a host.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Table 1. Deduced amino-acid sequence of semaphorin gene family.

Approximate position of enumerated peptide classes are indicated by parenthetical (a) through (o); semaphorin domains are bounded by arrows; G: grasshopper semaphorin I (SEQ ID NO:58), T: Tribolium semaphorin I (SEQ ID NO:64), D1: Drosophila semaphorin I (SEQ ID NO:60), D2: Drosophila semaphorin II (SEQ ID NO:62), H3: Human semaphorin I (SEQ ID NO:54), V4: Vaccinia virus semaphorin IV (SEQ ID NO:56), V5: Variola virus (human small pox) semaphorin IV (SEQ ID NO:66); small case residues: conserved residues; underline: signal sequence; solid bar: transmembrane domain; double dashes: immunoglobulin domain.

```
G                        MRAALVAVAALLWVVALHAAAWVNDVSPKMYVQFGEERVQR
T                          MVVKILVWSICLIALCHAWMPDSSSKLINHFKSVESKS
D1
D2 MSLLQLSPLLALLLLLCSSVSETAADYENTWNFYYERPCCTGNDQGNNNYGKHGADHVRE
H3                  MGWLTRIVCLFWGVLLTARANYQNGKNNVPRLKLSYKEMLESNNVIT
V4                                    MMVLLHAVYSIVFVDVIIIKVQRYINDI

    ↓
G  f-Lg--nESHKDHfKLLeKDHNSlLvga--rNI---VYnISLRDLTEFTEQRiEwHSsGAHRELcY
T  fT-g--nATFPDHfIVLNQDETSILvgG--rNR---VYnLSIFDLSERKGGRiDwPSsDAHGQLcI
D1
D2 fNCgKLY---YRTfHMNeDRDT-lYvgaMDrVF---RVnLQNISSSNCNRDAiNLEPTRDDVVScV
H3 fN-gLAnSSSYHTfLLDeERSR-lYvgaKDHIFSFDLVnI-------KDFQKiVwPVsYTRRDEcK
V4 LTLDIFYLFKKMIPLLFILFYFANGIEWHKFETSEEIISTYLLDDV------LYTGVNGAVYTFSN

                                  A
                        F         V
          DDCQN-YIR(a)            ICGTN(b)
G  LkgkS-Eddcqn-yir--VlAKIDDDrVLIcgtnaYKpLcRHyALKd-----GDyVVeKEYEgRg----
T  LkgkT-Dddcqn-yir--ilYSSEPGKlVIcgtnSYKpLcRTyAFKE-----GKyLVeKEVEgIg----
D1       Eddcqn-yir--iMVVPSPGrlFvcgtnSFRpMcNTyIISd-----SNyTLeATKNgQA----
D2 SkgkSQIFdcKnHViQ--SMDQGD--rlYvcgtnaHNpKDYViYANL-----THLPRSEYVIgVgLGIA
H3 WAgkDILKEcAn-FiK--VlKAYNQTHlYAcgtGaFHpIcTYiEIGHPEDNIFKLENSHFENgRg----
V4 NkLNKTGLTNNn-yiTTSiKVEDADKDTLvcgtnNGNpKcWKiDGSd--------------------

S F                                  A
CPYDP(c)                             TVADFSG(d)
```

```
G  LcpFdpDh---------nstAIYSEgQ-------lysAtvadfsgTdpLiyrGpl------------
T  LcpyNpEh---------nstsVSYNgQ-------lFsAtvadfsgGdpLiyrEpQ------------
D1 VcpydpRh---------nstsVLADNE-------lysgtvadfsgSdpIiyrEpl------------
D2 KcpydpLD---------nstAIYVENGNPGGLPGlysgtNaEfTKAdTViFrTDlYNTSAKRLEYKF
H3 KSpydpKL---------LTAsLLIDgE-------lysgtAadfMgRdFAiFrT-lGHHHPIRTEQHD
V4 ----dpKhRGRGYAPYQnsKVTIISHNGcYLSDINIsKEGIKRWRRFDGPcGYDl------------
V5                                                      MIYl------------

                                        N
              LNAPNFV(e)      (f)FFFRETA EYINCGK(g)       (h)DKGG
G  -rteRSdLkQ-lnapnfv-NTMEyNdFIFfffretaveyincgkaiysrvarvckHdkgg
T  -rteLSdLkQ-lnapnfv-NsVAygdYIFffYretaveyMncgkViysrvarvckDdkgg
D1 -QteQYdSLS-lnapnfv-SsFtQgdFvyfffretaveFincgkaiysrvarvckWdkgg
D2 KrtLKYdSkW-lDKpnfv-GsFDIgEYvyfffretaveyincgkaVysriarvckKdVgg
H3 -SRWLNdpkF-ISaHLISESdNPEDdkvyfffreNaIDGEHSgkaTHAriGQIckndFgg
V4 YTADNVIpkDGlRGA-fvDKdGty-dkvyILfTDtIG-SKRIVkIPy--iaQMcLndEgg
V5 YTADNVIpkDGlQGA-fvDKdGty-dkvyILfTVtIG-SKRIVkIPy--iaQMcLndECg

           SSY(i)     V
   PH      WTTFLKAR NCSIPG(j)
G  phQF-GDrwtsflkSrlncsVpgDypfyf---neiqs---tsdIIegNyGGQVEkliygv
T  phQ-SRDrwtsflkarlncsipgEypfyf---Deiqs---tsdIvegRyNsDDskIiygI
D1 phRF-RNrwtsflkSrlncsipgDypfyf---neiqs---AsNLvegQyGsMSskliygv
D2 KNLl-AhNwAtYlkarlncsiSgEFpfyf---neiqs---VYQL-----PsDKsRF-FAT
H3 -hRSLVNKwttflkarlIcsVpgPNGIDTHf-DeLq------dVFLMNFKDPKNPVVygv
V4 pSSlSShrwStflkVElEcDiDgRSYRQIIHSRTiKTDNDtILYvF--FDsPYsk-----
V5 pSSlSShrwStLlkVElEcDiDgRSYSQINHSKTiKQIMIRYYMYSLIVLFQVRIMYLFY

                                        V
            GSAVC(k)                    NSNWLPV(l)     PRPGTCVND(m)
G  fttpVnSiGgsavcafsmKSiLESfDgPfkeqETMnsnwlAvPSLKvpeprpgQcvndsr
T  LttpVnAiGgsaIcayQmAdiLRVfEgSfkHqETInsnwlpvPQNLvpeprpgQcvRdsr
D1 fNtpSnSiPgsavcafALQdiADTfEgQfkeqTGInsnwlpvNNAKvpDprpgScHndsr
D2 fttSTnGLIgsavcSfHINEiQAAfNgKfkeqSSSnsAwlpvLNSRvpeprpgTcvndTS
H3 fttSSnIFKgsavcMysmSdVRRVfLgPYAHRDGPnYQwVp-YQGRvpYprpgTc--PsK
V4 ----------saLcTysmNTiKQSfSTSKLeg----------YTKQLpSpApgIcLPAGK
V5 EYH

G  ---------TlpdVSVnfV-kShTlmdEAvpaFfTRpilIrIslQyrftKiAvdQqvRtPDgKAYdvLf
T  ---------IlpdKNVnfi-kThSlmED-vpaLfGKpVlVrVslQyrftAiTvdPqvKtINNQYLdvLY
D1 ---------AlpdPTLnfi-kThSlmdENvpaFfSQpilVrTsTIyrftQiAvdAqIKtPGgKTYdvIf
D2 ---------NlpdTVLnfi-RShPlmdKAvNHEHnNpVYYKRDlVFTK-LVVDKIRIDILNQEYI-vYY
H3 TFGGFDSTKDlpdDVITfA-rshPAmYNPvFPMNnRpiVIKTDVNyQftQiVvd-RvDAEDgQY-dvMf
V4 ---------VVpHTTFEViEKYNVlDdIIKp-LSnQpiFEGPSGVKWFDIKEKENEHREYRIYFIKENS

G  igtddgkvIkALnSAsFDSSDTvDSvVIeeLQvLPPGVpVKnlYVvr-------Mdg--d
T  igtddgkvLkAvnIPKRHAKALLYRKYRTSVHPHGA--pVKQlKIAP------------G
D1 VgtdHgkIIkSvnAEsADSADKvTSvVIeeIDvLTKSEpIRnlEIvrTMQYDQPKdgSYd
D2 VgtNLgRIYkIvnGEsLSKLLDIFEvAPNeAIQVMEISQTR--------------------
H3 igtdVgTvLkVvSIPKETWY-DLEEvLLeeMTvFREPTAISA------------MELSTK
V4 iYSFdTkSKQTRSSQVDARLFSvMVTSKPLFIADIGIGVGMPQMKKILKM*

                                        DPYCAWD(n)
G  dSklVVvSdDEiLAiKlhrcGSdkItNcRecvSlqdpycawdNVELKcTAVgSpDwSAG
T  YGkVVVvGKDEiRLANlNHcAS-k-tRcKDcvElqdpHcawdAKQNLcVSIDTVTSY--
D1 dGklIivTdSQVVAiQlhrcHNdkItScSecvAlqdpycawdKIAGKcRSHgApRw-LE
D2 -KSlYiGTdHRiKQiDlAMc-NRRYDNcFRcv--RdpycGwdKEANTcRPY--------
H3 QQQlYiGSTAGVAQLPlhrcDIYG-KAcAecCLARdpycawd--GSAcS---RYFPTAK

                                        ↓
G  KrRFIqNISLgEH-KAcGGRPQTEIVASPVPTQPTTKSSGDPVHSIHQAEFEpeiDNEiVI
T  -rFLIqdvVRgDD-NKcWsPQTDKkTVIKNKPSEVENEITNSIDEKDLDsSdpLiKTGLdD
```

```
D  ENYFYqNvATgQH-AAcPsGKINSkDANAGEGKGFRNDMDLLDSRRQ--sKdQeiIDNidK
D2 ELDLLqdvANETS-DIcDsSVLKKk
H3 RrTRRqdIRNgDPLTHcSDLHDNHH

G  GVddSNVIPNTLAEINHAGSKLPSSQEKlPiytaetlTiaIvTSCLGAlVvgfIsgFLFS
T  DSdcDPVSENSIGGcAV--------RQQlViytaGtlHiVvvVVsiVGlFSWLYsgLSVF
D  NFEdD----------------------IINAQytVetlVMavLAGsiFSlLvgfFTgYFCG

G  rrcRGEDYTDMpFpdQRHQLNRLTEAGlNADsPYLPPCANnkAAInlvLNv-----PpkN
T  AKFHSd----SQypEAPFIEQHNHLERlsANQTGYLTPRAnk-AVnlvvKvSSSTPRpkK
D  rrcHKdEDDNLpypdTEYEYFEQRQNVNsFPsSCRIQQEPKLLPQVEEvTYAEPVLLpQP

                        KKTYI(o)
G  AnGKNANsSAENKP----IQkktyi*
T  DnLDVSKDLNIASDGTLQKIkktyi*
D  PPPNKMHsPKNTLRKPPMHQMHQGPNSETLFQFHVTATTPSSRIVVATTSEHCVPTR*

D2                  -----IVVTyg---QsVHlGcFVkIPEVlKNEQvTwYHHSKDKG
H3                  GHSPEERIIygVENSsTFlEcSPkSQRAl----vYwQFQRRNEE
                                ============================

D2 rYeIRYSPTKYiETtERglVVVsVNEAdGgRyDchLGGSLLcSYNITVDAHRcTPPNKSN
H3 rKeE-IRVDDHiIRtDQglLLRsLQQKdSgNyLchAVEHGFIQTLLKVTLEVIDTEHLEE
   ======================================

D2 DYQKIYSDWcHEFEKYKTAMKSWEKKQGQcSTRQNFScNQHPNEIFRKPNV*
H3 LLHKDDDGDGSKTKEMSNSMTPSQKVWYRDFMQLINHPNLNTMDEFcEQVWKRDRKQRRQ

H3 RPGHTPGNSNKHLQENKKGRNRRTHEFERAPRSV*
```